

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:43 ; Search time 91.75 Seconds

(without alignments)  
219.987 Million cpi; updates/sec

Title: US-09-331-631A-5

Perfect score: 3326  
Sequence: 1 00MOLETSQOMRCVSCQDK.....SPRSTKQQQPLVSIIDFVG 625

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Query length	DB ID	Description
1	1074.5	32.3	588	1	VCLB_GOSHI
2	1063.5	32.0	605	1	VCLB_GOSHI
3	900	27.1	573	1	GLB1_MAIZE
4	863.5	26.0	626	1	AH12_ARAHY
5	861	25.9	614	1	AH11_ARAHY
6	859.5	25.8	605	1	GLCA_SOYBN
7	854.5	25.7	639	1	GLCX_SOYBN
8	833	25.0	571	1	CVCA_PEA
9	794	23.9	524	1	SBP_SOYBN
10	759.5	22.8	459	1	VCLC_PEA
11	749	22.5	439	1	GLCB_SOYBN
12	713.5	21.5	463	1	VCL_VICFA
13	712.5	21.4	410	1	VCLB_PEA
14	692.5	20.8	445	1	CANA_CANEN
15	685.5	20.6	445	1	CANA_CANGL
16	563	16.9	436	1	PHSA_PHAVU
17	552	16.6	421	1	PHSR_PHAVU
18	493	14.8	386	1	CVCB_PEA
19	414	12.4	275	1	VCLB_PEA
20	216	6.5	1888	1	TRHY_HUMAN
21	166	5.9	1407	1	TRHY_RABIT
22	193	5.8	499	1	GL02_ORYSA
23	192.5	5.8	124	1	VCLU_PEA
24	192.5	5.8	500	1	GLU5_ORYSA
25	191	5.7	499	1	GLU8_ORYSA
26	185.5	5.6	1549	1	TRHY_SHEEP
27	185	5.6	518	1	SSG2_AVEA
28	184	5.5	499	1	GL04_ORYSA
29	184	5.5	498	1	SSG1_AVEA
30	180.5	5.4	495	1	GLU2_ORYSA
31	180.5	5.4	499	1	GU11_ORYSA
32	180	5.4	496	1	GLUC_ORYSA
33	176.5	5.3	1023	1	GLT_DROME

34	176	5.3	544	1	INVO_AOTTR	P24708 actus trivi
35	170.5	5.1	338	1	LEGB_PEA	P14594 pisum sativ
36	169.5	5.1	585	1	INVO_HUMAN	P07476 homo sapien
37	169	5.1	522	1	INVO_HYLLA	P17941 hylobates l
38	166.5	5.0	496	1	GLU3_ORYSA	009151 oryza sativ
39	165	5.0	708	1	GBF_DICDI	P36417 dictyostell
40	164	4.9	560	1	INVO_PANPA	P14591 pan paniscu
41	162.5	4.9	471	1	RU17_XENLA	P09406 xenopus lae
42	162	4.9	516	1	LEGB_GOSHI	P09800 gossypium h
43	160	4.8	835	1	INVO_PONPY	P14708 pongo pygma
44	157.5	4.7	493	1	INVO_SAGOE	P24712 saguinus oe
45	157.5	4.7	877	1	INCE_CHICK	P53352 gallus gall

## ALIGNMENTS

RESULT	ID	VLB_GOSHI	STANDARD	PRT	588 AA.
AC	P09801				
DT	01-MAR-1989	(Rel. 10, Created)			
DT	01-MAR-1989	(Rel. 10, Last sequence update)			
DT	15-JUL-1999	(Rel. 38, Last annotation update)			
DE	VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).				
OS	Gossypium hirsutum (Upland cotton).				
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.				
RN	(1)				
RA	SEQUENCE FROM N.A.				
RP	Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;				
RT	"Developmental biochemistry of cottonseed embryogenesis and germination. XVIII. cDNA and amino acid sequences of the members of the storage protein families."				
RL	Plant Mol. Biol. 7:475-489(1986).				
CC	-1- FUNCTION: SEED STORAGE PROTEIN.				
CC	-1- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN BODIES.				
CC	-1- SIMILARITY: NO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN, CONVICILIN, CONGLICILIN, ETC.).				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	EMBL: M16891; AAA33071.1; -				
CC	DR PIR: A30838; FMCNAB.				
CC	DR HSSP: P50477; ICAX.				
DR	INTERPRO: IPR00113; -				
DR	PFAM: PF00546; Seedstore_7s; 1.				
CC	Seed storage protein; Signal.				
FT	SIGNAL	1	25		
FT	CHAIN	26	588		
SO	SEQUENCE	588 AA; 69729 MW; 63699829ABADEB CRC64;			

Query Match 32.3%; Score 1074.5; DB 1; Length 588;  
Best local Similarity 38.2%; Pred. No. 1.2e-59;  
Matches 229; Conservative 115; Mismatches 181; Indels 75; Gaps 13;

QY	35	EDPQTCQCCQRCRQOESDPQOQYQCRCKEICEEENROR--DPQOYEQCQKRC	92
DB	35	DDEPKRYEDCRRCSEWDTGKQEQCCQSCSKSQQYGEKQDQQRHREDDPQRRYEECCQCC	94
QY	93	ORRETEPRHMQICQRCSERYEKEKKQKRYEEQOREDEKYEEMKKCDNKRPOQOE	152
DB	95	--RQEEKRPQCCQRCRLRFQEEQOO-----SQHQ 123	

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@sib-sib.ch).  
CC -----  
DR EMBL; M19378; AAA33069.1; .  
DR PIR; S06398; S06398.  
DR HSSP; P50477; ICAX.  
DR INTERPRO; IPR001113; .  
DR PfAM; PF00546; SeedsStore\_7s; 1.  
DR Seed storage protein, Signal.  
KW SIGNAL  
FT 23

FT	CHAIN	24	605	VICILIN GC27-A
SO	SEQUENCE	605 AA:	71049 MW:	C9DB9371C976953B CnC64:
	Query Match	32.0%;	Score 1063.5;	DB 1; Length 605;
	Best Local Similarity	38.0%;	Pred. No. 5,9e-59;	
	Matches 232;	Conservative 121;	Mismatches 195;	Indels 63; Gaps
OY	NOEPOTECCOCCORRCQOQSDPDRQOQYORCKELCEEDEEYNRORPOOYEDCCORC	92		
DB	33 SEDPQORECECKRCKOLETGTOTEDCKEDNSEYOLKEEQORDE-DPQRYQDCRHC	91		
OY	QRRLETER-HMOICQOCERCERYEKERKQOKRYEEOQRDEEYKYEEMKEGDKHDPQR	151		
DB	92 QOEBRLRLPH--CEQSCREYQYK-----QQQQQPD-----K	120		
OY	152 EYEDCRKHCEQD--PRLYQOCORQOEDQORQHGKRGDLMNQ-R-GSGRYEDEGEKSD	208		
DB	121 QFKCCQOCRCQOEDRPKQOCVCEKREYOEDPYMKGERENKWEDEEEDSDQOORN	180		
OY	209 NPYYEDRSLSRTRETEEGHISLVENFYGSKLLALRYRLVLLLEAPNPAVFLPTHDA	268		
DB	181 NPYYFHRSQCERRERHNGFRLQFRADKHNHLRQINFRAILLLEAPNPFVLRPHCDA	240		
OY	269 DALLVIGRGALKMILHRDNRESYMLECGDVTIRPAGTTFYILINDNNEBRLIAKFLQTI	328		
DB	241 EKIVVYNGRGTVFYVYHENKESYVNVPGVVYRIPAGSTVYLANODNBEKLTIAVLAHrv	300		
OY	329 STPGQKEPFGAGQONDEPLSTFSKEILLLEALNPTOTELRVLG-----QOREGVII	381		
DB	301 NNPOQOFKFPFAGQENFQSLTRFSKEILLEAVFNTRSEDLDEPCGROSHRQOQSGNFR	360		
OY	382 RASQEQIRELTRDDESRMRHIRGGESSRG-PYULFNKRPRLYSNKYQAYEVKPEDEY-R	439		
DB	361 KASQEQIRALSGQATSPR-----GKSGEGYAFNLLSPTPSNONGRFGYCAPRNFQ	413		
OY	440 OLOQMDVSVFLANTQSSMMGPFNTRSTKYVVVYVVASGADVEMACPHLSGRIGSGGSKR	499		
DB	414 QLRVEDSSVAAFETKNSLIVPHYNSKATFYVLVYEGNHWEMVCPHLS-ROSSDWSRE	472		
OY	500 HEEEEE-----HYGEVBARLSKREALIVLAGHPVNVSSGNNELLLFAFGI-NAQNN	551		
DB	473 EEEQDEQEVERRSQYKRVAAQJLSTGNLFYPVAGHGVTFVASONEDLGLLGGLGNQDN	532		
OY	552 HENLAGRERNVLOQIQPMELAPASKEVEBELFNSODESIFPPRGPRQHQOQSPRSTK	611		
DB	533 KRIVPAKTNVY-RQMDROAKELAFGESRLYDEVFNNNPOESTYFVSGHRRGFDER-RG	590		
OY	612 QOQPLVSLDP 622			
DB	591 SNNPLSEPLDF 601			
RESULT 3				
GLBL_MAIZE	STANDARD;	PRT;	573 AA.	
AC	P15590;			
DT	01-APR-1990 (Rel. 14, Created)			
DT	01-AUG-1990 (Rel. 15, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	GLOBULIN-1 S ALLELE PRECURSOR (GLB1-S) (7S-Like).			
GN	GLB1.			
OS	Zea mays (Maize).			
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
OC	Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-CV. INBRED LINE VA26;			
RA	Belanger F.C., Kriz A.L.;			
RT	"Molecular characterization of the major maize embryo globulin encoded			
RT	by the Glb1 gene".			
RL	Plant Physiol. 91:636-643(1989).			
RN	[1]			



[illegible]

RESULT	5			
AC	AH11_ARAHY	STANDARD:	PRT:	614 AA.
AC	P43237			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	ALLERGEN ARA H 1, CLONE p17 (ARA H 1).			
OS	Arachis hypogaea (Peanut).			
CC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons, core eudicots; Rosidae; eurosids I;			
CC	Fabales; Fabaceae; Papilionoideae; Arachis.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. FLORUNNER;			
RX	MEDLINE=96013631; Pubmed=7560062;			
RA	Burks A.W., Cockrell G., Stanley J.S., Helm R.M., Bannon G.A.:			
RT	"Recombinant peanut allergen Ara h 1 expression and IgE binding in			
RL	patients with peanut hypersensitivity."			
J.	Clin. Invest. 96:1715-1721(1995).			
CC	- I - SIMILARITY: TO 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,			
CC	CONVICILIN, CONGLYCININ, ETC.).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce@lsb-sib.ch">http://www.isb-sib.ch/announce@lsb-sib.ch</a> ).			
CC	or send an email to <a href="mailto:license@lsb-sib.ch">license@lsb-sib.ch</a> ).			
CC	-----			
DR	EMBL: L38853; AAA60336.1; -			
DR	HSSP: P50477; ICAM			
DR	INTERPRO: IPR001113; -			
DR	PFAM: PF00546; Seedstore_7s; 1.			
DR	Allergen.			
QO	SEQUENCE 614 AA: 70283 MW; 1DDACF217EEC5F31 CRC64;			

Query Match	25.98;	Score 861;	DB 1;	Length 614;
Best Local Similarity	34.58;	Pred. No. 2.1e-46;		
Matches 220;	Conservative 114;	Mismatches 205;	Indels 98;	Gaps 23

	QY	29	SKYNODDPOTEC--QOCORRCROEESPPOOYOVCARCKICEEBEEFYNNRQPOOYEG	87
	Db	27	SPYKTEPNP---CQCRLOQS-QDEPDCLKOKACESCCTKL-----EX-----DPCRYVD-	72
	QY	88	CQKCQRREPRIMQICQCRCERRYEKKRKQOORYEDQREDEEKYEERMEKGD--NK	145
	b	73	--TGATNRHP-----GRTGRGPGDV-----DDIKRQPRREGEWGWF	111

```

Oy      146 KDPOREVEDRHRCEODEPLOYOCORPOEO-----POHRCGLDMNFORGSGHYEE 201
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      112 AEPRRERE-----DMQRPRED---RRSIQOPKRIREDGRGE-----QEWGTPESEV 159
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      202 GEEKOSNDPYFEDERSLSTRETEEGHISYLENYFGSKLLRALKNRYLVLEANPAFY 261
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      160 REETSRNNPEFFPSRSTRGMONGRLVLQREFDORSKOFOUONNRIVOTIARPPTLV 219
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      262 LPTHLDAILLYVGGRALMKIRDRRESYNLECGRVITHPAGTFYLLRNPNNEBTH 321
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      220 LPKHADADNIIYIOGGATTVVANGNNKRSFNDEGALAIIPSCFIYTILNRIDNONLRV 279
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      322 AKPLTISTPGQYKEFPAGONEPYLSFTFSKEILEALNOTERLGVL----- 372
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      280 AKISMVPVTPOGFEDFFPASSRDOSXYLOGFSRNTLEAANAENELRNVILLEENGGEQ 339
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      373 ---GOOR-----EGYIHASOEQRBELTRDSESRMHIRRGSSS---KGPNYLFRK 419
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      340 EERGORKRSTRSSNEGVIYAVKSHKEHOELTKHAKSYS----KKGSPEEDITNPINLRDG 395
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      420 RPLYSNKRGQAYEVKRPEDYR-QLODMOVSYFIANTIGWSMNGCFEPPNRTSRVVVASGEA 478
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      396 EPDLSNNGRGLEHVKKDPKKNPQLDDMLMTLCVEIKEGALMLPHFNKSAMVIVYVKMGTS 455
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      479 DYEMACPHLSRHRGGRGGKHHHEEEVHYE---OVR---ARLSKREALVYLAGRPFVF 531
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      456 NLELVAARKEOQGGRREDEWEEDDEEBEGESNRERYARTARLKEDVDIMPAAHVAI 515
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      532 VSSGNENIELPAFINOMNIENFLAGRERVLOIQEPOMALEAFASRKREVEELEFNSOD 591
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      516 NASSLRHLL--GFGIMENNRRIFELADDKDQVLDQIEKOKADLAFLPSSGOYEKLTIKNOR 573
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      592 ESIFPG-PROHQOOSPASTKQOO-----PLVSIL 620
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      574 BSHFVSARPOSQSPSEKEDEEENOGKGPULSIL 610

```

RESULT	6
ID	GLCA_SOYBN
AC	GLCA_SOYBN
DT	13916,
DT	01-JAN-1990 (Rel. 13, Created)
DT	01-APR-1990 (Rel. 14, Last sequence update)
DT	15-DEC-1998 (Rel. 37, Last annotation update)
DE	BETA-CONGLYCININ, ALPHA CHAIN PRECURSOR.
OS	Glycine max (Soybean).
OC	Eukaryota, Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC	Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eurosids I;
OC	Fabales; Fabaceae; Papilionoideae; Glycine.
RP	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=COYLEDON;
RX	MEDLINE=91355860; Pubmed=2103438;
RA	Sebastian F.L., Farrel L.B., Schuler M.A., Beachy R.N.;
RT	"Complete sequence of a cDNA of alpha subunit of soybean beta-
RT	conglycinin";
RL	Plant Mol. Biol. 15:197-201(1990).
CC	-1- FUNCTION: SEED STORAGE PROTEIN. ACCUMULATES DURING SEED
CC	DEVELOPMENT AND IS HYDROLYZED AFTER GERMINATION TO PROVIDE A
CC	CARBON AND NITROGEN SOURCE FOR THE DEVELOPING SEEDLING.
CC	-1- SUBUNIT: THE ALPHA-, ALPHA-, AND BETA-SUBUNITS ASSOCIATE IN
CC	VARIOUS COMBINATIONS TO FORM TRIMERIC PROTEINS.
CC	-1- SUBCELLULAR LOCATIONS: EMBRYO AXIS, AND COYLEDONARY MEMBRANE-BOUND
CC	VACUOLAR PROTEIN BODIES.
CC	-1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC	CONVICILIN, CONGLYCININ, ETC.).
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ).

```

CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X17698; CAA35691.1; -.
DR PIR: S14681; FMSYBA.
DR HSSP: P50477; ICAM.
DR INTERPRO: IPR001113; -.
DR PFM: PF00546; Seedstore_7s; 1.
KM Seed storage protein; Signal; Glycoprotein; Multigene family.
FT SIGNAL 1 22
FT PROPEP 23 62
FT CHAIN 63 605 BETA-CONGLYCININ, ALPHA CHAIN.
FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 517 517 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 605 AA: 70293 MW: CBBAA30506BBBC57 CRC64;

Query Match 25.8%; Score 859.5; DB 1; Length 605;
Best Local Similarity 32.1%; Pred. No. 2.6e-46;
Matches 203; Conservative 133; Mismatches 197; Indels 99; Gaps 18;

QY 31 YDNOEDPQTECCOQRCRQOESDPQOQYQORRC-----KEICEEE-----EYNR 77
| : : : : : | : : : : : | : : : : : | : : : : : |
Dy 27 YMEKENPKH--NKCLQSC-NSESDYRNQACHARNLKVKECECECEIIPRPRAPRP 83
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 78 QRPQOQYEQCQRCQQR-----RETEPRHMQICQRCERRY-EKEKKQOQRYEQQRE 130
| : : : : : | : : : : : | : : : : : | : : : : : |
Dy 84 EREPQOQGEKEDEDEDPRIPPRPPQOPEEHEQEEOEWPKEKREKSEDEED 143
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 131 DEKKEERMEKGNKRDPOQREYEDCRHCEQOEPRLQYQCCRCQOQOQNHGSGDLN 190
| : : : : : | : : : : : | : : : : : | : : : : : |
Dy 144 EDEEODER--QFPFPRPHQKE-----ERNEEED-----BEQOQES----- 179
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 191 PORGSGRYEEGEQK-----SDNPYFDEERLSSTRFTEEGHISYLENFYKSKLLRAL 245
| : : : : : | : : : : : | : : : : : | : : : : : |
Dy 180 -----EESDSELRHKKNNKPNFEGSNRFETLFKNQYGRIVLQNRNQSPQLQNL 230
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 246 KNRVLVLEENPAFVLPFLHDALLVYGGAALKMIRHNRSEYNLECGDVITIPAG 305
| : : : : : | : : : : : | : : : : : | : : : : : |
Dy 231 RDRILEFNSKPTLLLPNADADYIVILNGTAIISLVNNDRDSYRLQSGDALRVPSS 290
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 306 TPFFYLNRDNNERLHAKFLQITSTPGQYKEFPFAGQNDPEPLSPFSKEILEALNTQT 365
| : : : : : | : : : : : | : : : : : | : : : : : |
Dy 291 TTYTYVNPNNENLKRITLTAIPNNKRGRESFSLSTEQASTLGQFSKRILEASTDYK 350
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 366 ERLRGVL-----GOQR--EGVITIRASQEOIRELTRDSESRWMIIRGCESSRGY 414
| : : : : : | : : : : : | : : : : : | : : : : : |
Dy 351 EEINKVLFSREGEQOQGEORLOESYIVELSKQITRAKSSSKRTI-----SSDKPF 406
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 415 NLFNKRPLYSNKGAYEVKPEDEYRQLODMQSVFTANTTQSGMGPFTNTSTKYVVA 474
| : : : : : | : : : : : | : : : : : | : : : : : |
Dy 407 NLSRDPISYNNKLGKFFETIPKKNPQLRDLIDFLSTIVDMNEGALLPHNSKAIVILIN 466
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 475 SGADYEMACPHLSGRHGGGCKRHEEPEEYEQ-----VRATLSKREALIVLAGHP 528
| : : : : : | : : : : : | : : : : : | : : : : : |
Dy 467 EGDANIELV-----GLKEQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 513
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 529 VVFVSSGNENLLFAPGINAONNHNENFLAGRENNVLOQIEPQAMETAFASKEVELEFN 588
| : : : : : | : : : : : | : : : : : | : : : : : |
Dy 514 VVV--NATSNLNFALGICNAENNRNFLAGSDQNVISQILPSQVQELAFGSGQAVKLLK 571
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 589 SQDESTFFPGPRHOQOQSPRSTKQOQPLYSIL 620
| : : : : : | : : : : : | : : : : : | : : : : : |
Dy 572 NQRESYFYDAOPKKKEGNKGR--GPISSIL 601
| : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 7
GLCX_SQOYBN STANDARD; PRT; 639 AA.
AC P11827;
DT 01-OCT-1989 (rel. 12, Created)
DT 01-OCT-1989 (rel. 12, Last sequence update)
DT 01-AUG-1992 (rel. 23, Last annotation update)
DE BETA-CONGLYCININ, ALPHA CHAIN PRECURSOR.

```

```

GN CG-1.
OS Glycine max (soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids I;
OC Fabales; Fabaceae; Papilionoideae; Glycine.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86250867; PubMed=3013879;
RA Doyle J.J., Schuler M.A., Godelte W.D., Zenger V., Beachy R.N.,
RA Slighmon J.L.;
RT "The glycosylated seed storage proteins of Glycine max and Phaseolus
vulgaris. Structural homologs of genes and proteins.";
RL J. Biol. Chem. 261:9228-9238(1986).
RN [2]
RP SEQUENCE OF 340-639 FROM N.A.
RX MEDLINE=83143289; PubMed=6897678;
RA Schuler M.A., Ladin B.F., Pollaco J.C., Freyer G., Beachy R.N.;
RT "Structural sequences are conserved in the genes coding for the
alpha, alpha' and beta-subunits of the soybean 7S seed storage
protein.";
RL Nucleic Acids Res. 10:8245-8261(1982).
CC -I- FUNCTION: SEED STORAGE PROTEIN. ACCUMULATES DURING SEED
DEVELOPMENT AND IS HYDROLYZED AFTER GERMINATION TO PROVIDE A
CARBON AND NITROGEN SOURCE FOR THE DEVELOPING SEEDLING.
CC -I- SUBUNIT: THE ALPHA', ALPHA', AND BETA-SUBUNITS ASSOCIATE IN
VARIOUS COMBINATIONS TO FORM TRIMERIC PROTEINS.
CC -I- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
BODIES.
CC -I- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CONVULGIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M13759; AAB01374.1; -.
DR EMBL: J01280; -. NOT_ANNOTATED_CDS.
DR PIR: B24810; B24810.
DR HSSP: P02853; 2PHL.
DR INTERPRO: IPR001113; -.
DR PFM: PF00546; Seedstore_7s; 2.
KM Seed storage protein; Signal; Glycoprotein; Multigene family.
FT SIGNAL 1 22
FT PROPEP 23 62
FT CHAIN 63 639 BETA-CONGLYCININ, ALPHA CHAIN.
FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 551 551 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 543 543 P -> L (IN REF. 2).
FT CONFLICT 549 549 M -> V (IN REF. 2).
FT CONFLICT 608 608 S -> T (IN REF. 2).
SQ SEQUENCE 639 AA: 74325 MW: 469BF24C79651E3F CRC64;

Query Match 25.7%; Score 854.5; DB 1; Length 639;
Best Local Similarity 32.0%; Pred. No. 5.6e-46;
Matches 203; Conservative 141; Mismatches 215; Indels 79; Gaps 19;

QY 28 WSKYNOEDPQTECCOQRCRQOESDPQOQYQORCKEI-CEEEEEVN-----FOR 79
| : : : : : | : : : : : | : : : : : | : : : : : |
Dy 28 WEKQNPShN-----KCLRSC-NSEKDSYRNQACHARNLKVREBECEDEQOITPRAPQ 80
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 80 DPQOQYEQCQRCQQRRETEPRHMQICQRCERRYEKKRKQOQRYEQQRED-----E 132
| : : : : : | : : : : : | : : : : : | : : : : : |
Dy 81 HPEREQOQGEKEDEDEGEOPRPFPPRPR--QPHQEEHEQOKEHEHMRKEKHGKQSE 138
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 133 EKYEREMKGDGDKRDPQOQREYEDCRHCEQOEPRLQYQCCRCQOQOQOQOQOQOQOQOQOQ 191
| : : : : : | : : : : : | : : : : : | : : : : : |
Dy 139 EEDER--EHPRPHQOKEE--KH-EWQHKQKHGKSEEEDEDEDEDEQDQKESQ 192
| : : : : : | : : : : : | : : : : : | : : : : : |

```





```

OY 269 DALLVIGRGALKMIHNDNRESYULCEGDVIRIPAGTFFYLINBNRCHIAKFLQTI 328
Db 94 DILLVLVSGKATLILYKPKDRNSFNLEBDRTIKLPAGTATLVYVNRDDBELKVLDAIPV 153
OY 329 STPGQYKEFFPAGGONPEBYLSTFKEKLEALINTOTERLRCVL----- 372
Db 154 NRPGLQDSPLSLGNOONQVYLSGFSKNLEASFNFDYEIEKVLLEHBEKETOHRSJKD 213
OY 373 ---GQOREGVIIIRASOEQIRELTRDSDSRKRRHIRGCGSSGPPYNLFKKRLYSNKQGO 429
Db 214 KRQOQOEENVLYIKLSRGOYEELSKNAKSTST---KKSYSSESEPNLRGRGPIYSNEFG 269
OY 430 AREVPEPDRLOLQDDVDSVFANTITOGSMGCFPFNTSRKTVKVVVNSGSDVEMACPHLSG 489
Db 270 FEETPEPKRPQLODDIDIPVNSVEITEGSLILPHYSKRAIVATYVNGKDFELVQORNN 329
OY 490 RHGGRGGGRHREEEBEVHYEYOV-----RARLSKREALIVLAGHPVYFVSSGNENLLF 542
Db 330 QOEOR---KEDDEEEEGEGEELINKQOVYKAKLSQDVIVIPAGHPAVKASSNIDL-- 384
OY 543 AGCIAMONNHHEFLAGRENVYLOQLEPQAMELAFPAASKREVELELNSODESLFFPGPRQH 602
Db 385 GEGVIAENNQRNFGLAGDEBNVISQIRPYKELAFPGSAQEDVRILENOKSHFADAQPOQ 444
OY 603 QOQSPESTKQO 613
Db 445 RERGSRETRDR 455

RESULT 11
GLCB_SOYBN
ID GLCB SOYBN STANDARD: PRT: 439 AA.
AC P25974:
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE BETA-CONGLYCININ, BETA CHAIN PRECURSOR.
GN CG-4.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Glycine.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93005638: PubMed-2562562:
RA Harada J.J., Barker S.J., Goldberg R.B.:
RT "Soybean beta-conglycinin genes are clustered in several DNA regions
RT and are regulated by transcriptional and posttranscriptional
RT processes."
RL Plant Cell 1:415-425(1989).
CC -1- FUNCTION: SEED STORAGE PROTEIN. ACCUMULATES DURING SEED
CC DEVELOPMENT AND IS HYDROLYZED AFTER GERMINATION TO PROVIDE A
CC CARBON AND NITROGEN SOURCE FOR THE DEVELOPING SEEDLING.
CC -1- SUBUNIT: THE ALPHA-, ALPHA-, AND BETA-SUBUNITS ASSOCIATE IN
CC VARIOUS COMBINATIONS TO FORM TRIMERIC PROTEINS.
CC -1- SUBCELLULAR LOCATION: EMBRYO AXIS, AND CYTOSOL MEMBRANE-BOUND
CC VACUOLAR PROTEIN BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S44893: AAB23463.1; -.
DR PIR: J00969; FMSYCB.
DR HSSP: P02853; 2PHL.
DR INTERPRO: IPR001113; -.

```

DR	PFAM: PF00546; Seedstore7s; 1.	Glycoprotein; Multigene family.
KM	Seed storage protein: Signal:	
FT	SIGNAL	1 23
FT	PROPEP	24 25
FT	CHAIN	26 439
FT	CARBOHYD	351 351
SQ	SEQUENCE	439 AA; 50552 MW; DED8AA2A0776088B CRC64;
Query Match 22.5%; Score 749; DB 1; Length 439;		
Best Local Similarity 37.8%; Pred. No. 1.2e-35;		
Matches 162; Conservative 87; Mismatches 146; Indels 34; Gaps 7;		
QY	205 KOSDPYVF-DERSSTFRTEEGHISYLENPFYGRSKILRAIKNYRLVLEANPAVLP	263
DB	28 EDENNPFYFRSSNSQTLFENONVAIRLLQRPKNKSPOLENLRDRIYQFSKPTILLP	87
QY	264 THLDAAILLYIGRGALKMIIHNRRESYNLECGDVIRIPAGTFYLYINRDNERRLIAK	323
DB	88 HHADDFLLFYLSGRAITLTVNNDRDSDSYNLHPGDAQRIPTGYTYLVNPHDHLKLIK	147
QY	324 FLQITSTPGQVKEEFPAGGQNEPRLTSPFKELLEAINTQTERLGLV-----GOORE	377
DB	148 LAIPVKKGRFDPLFSLSTQAOOSTYLOGSHNLTSTSHSEFEELNRYLFGDEEERQOE	207
QY	378 GVIIASQEOIREFLRDSESRRMIRRGSGSSRPVYLFNKRPLYSNKYGAVYKPED	437
DB	208 GIVELSKQIQLNQLSRRAKSSSKRTI---SEDEPFLNRSNPIYSNPGFPEITPEK	263
QY	438 YRQLODMQVSVTIANITQGSMMKGFPPFNSTRKTVVYVAGADVEMACHLSRGCRGG	497
DB	264 NPQLRDLFIILSSVDYINCEGALLPHFNKAIVILVINGDNATIELV-----G	310
QY	498 -----KRHEEEHYHCOVARLRSKRAIVLAGHPVAVSSGNEMLLFAFGINAONN	551
DB	311 IREQOKQKQKEEPELEVGQYRRLSEDDVFIYIPAYP--FVYNATSNLFLAFGINAENN	368
QY	553 HENFLAGERNVLAQIEPQAMELAFASRKEVEELFNSQDESIFPPGPRHQOQSPSTK	611
DB	369 ORNPLAGEEDNVNRIEROVELAPFGSADVERLLKQKRESYFVDA--QPOKKEGSKG	426
QY	612 QOOPVLSII 620	
DB	427 RKGFPSTIL 435	
RESULT 12		
ID	VCL_VICFA	
AC	VCL_VICFA	STANDARD: PRT: 463 AA.
AC	PO8438:	
DT	01-AUG-1988 (Rel. 08, Created)	
DT	01-AUG-1988 (Rel. 08, Last sequence update)	
DT	01-JAN-1990 (Rel. 13, Last annotation update)	
DE	VICILIN PRECURSOR.	
OC	Vicia faba (Broad bean).	
OC	Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta:	
OC	Magnoliophyta: eudicotyledons, core eudicots; Rosidae; eurosids I;	
OC	Fabales; Fabaceae; Papilionoideae; Vicia.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=CV. MINOR.	
RX	MEDLINE=88096511; PubMed=3697075;	
RA	Weschke W., Baumelein H., Mobus U.;	
RT	"Nucleotide sequence of a field bean (Vicia faba L. var. minor) vicilin	
RT	gene.";	
RL	Nucleic Acids Res. 15:10065-10065(1987).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=CV. FRIBO.	
RX	MEDLINE=88067789; PubMed=3684610;	
RA	Bassener R., van Nong H., Jung R., Saalbach G., Muentz K.;	
RT	"The primary structure of the predominating vicilin storage protein	
RT	subunit from field bean seeds (Vicia faba L. var. minor cv. FRIBO).";	



RL Nucleic Acids Res. 15:9609-9609(1987).  
CC -1- FUNCTION: SEED STORAGE PROTEIN.  
CC -1- SUBCELLULAR LOCATION: CYTLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN  
CC BODIES.  
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASCOLIN, VICILIN,  
CC CONVICTILIN, CONGLICININ, ETC.).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Y00506; CAA68559.1; -  
DR EMBL: Y00462; CAA68525.1; -  
DR PIR: S06309; S06309.  
DR PIR: A27288; A27288.  
DR HSSP: P50477; ICAM.  
DR INTERPRO: IPR001113; -  
DR PFAM: PF00546; Seedstore\_7s; 1.  
KW Seed storage protein; Signal.  
FT SIGNAL 1 27  
FT CHAIN 28 463 VICILIN.  
FT CONFLICT 257 257 R -> K (IN REF. 2).  
FT CONFLICT 443 443 R -> Q (IN REF. 2).  
SQ SEQUENCE 463 AA; 52694 MW; 5CD09EFDE0D5E6B5 CRC64;  
  
Query Match 21.58; Score 713.5; DB 1; Length 463;  
Best Local Similarity 36.0%; Pred. No. 2.1e-37;  
Matches 158; Conservative 87; Mismatches 157; Indels 37; Gaps 6;  
  
QY 208 DNPYYEDERSLSTRFTEEGHISVLENFYGRSKLRALKNYRLVLLLEANPAFLPHLD 267  
DB 33 DNPVEFSNRPOQLTENENGHILRLQKFDQSKLENLQNRLLLEYSKPHPTFLPQOTD 92  
QY 268 ADAILLVIGRGALKMTHRDRESYNLECGDVIRIPAGTFYLLINRNNRLHIAFLQT 327  
DB 93 ADFLVLVLSGKALITVLLPDRNSFSLERCDITKLPAGTIGLVNRDDEDLRVLDVIP 152  
QY 328 ISIPGOYKEFFPAGGNPEPYLSTFESKLEALINTOTERLRYLQO----- 374  
DB 153 VNRGEPOSEFLSGNQNPSTLSGFSKNILEASFNTDYKLEKVLLEHGKERYHRGLK 212  
QY 375 -----GEGVIRASQEQIRELTRDDESRRMHIRGSGSSRGPNLFRKRPYSNKG 428  
DB 213 DRROGOEENVIVAKISKQJLELNKNAKSSS---KSTSESEFPNLRREPLYSNKFG 268  
QY 429 QAVEVKEPEDYROLQDMQVSVFIANTOGSMGPFENRSTKRVVAVASGEADVEMACPRLS 488  
DB 269 KFEFIFPKRNPQLODNLIFVNYVEINCSLLPLPHYNSRAIVTVYVNGCKDFELVQGRNE 328  
QY 489 GRHGGGCGKRHEEEVEVHYQVR-----ARLSKREAIIVLAGHPVYVSSGNENLL 541  
DB 329 NOOGLR---EYDEEKEGQGEIEIRKQYQNYRAKLSPGDVIVIPAGYFVAITAKSSNLLV- 384  
QY 542 FAFGINNONHNFENLAGRENVLOQIEPQAMELAFASRKEVELFNSQDESIFFPQPRQ 601  
DB 385 -GRCINAEENNRFTLAGEEDNVISQIHKPKELAFPGSAOEVDTLLENQKSHFANQPR 443  
QY 602 HQQSPRSTRKQOQPLVSIL 620  
DB 444 EREGSQEIKDH--LYSIL 460  
  
RESULT 13  
VCLB\_PEA STANDARD; PRT; 410 AA.  
AC P02854;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 01-JAN-1990 (Rel. 13, Last annotation update)  
DE PROVICTILIN PRECURSOR (TYPE B) (FRAGMENT).  
OS Pisum sativum (Garden pea).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
OC Fabales; Fabaceae; Papilionoideae; Pisum.  
RN [1]  
RP SEQUENCE FROM N.A. (CLONES PDUB7 AND PDUB4).  
RC STRAIN-CV. FELTHAM FIRST;  
RX MEDLINE=83220791; Pubmed=6687941;  
RA Lycett G.W., Delauney A.J., Gatehouse J.A., Gilroy J., Croy R.R.D.,  
RA Boulter D.;  
RT "The vicilin gene family of pea (Pisum sativum L.): a complete cDNA  
RT coding sequence for preprovictilin.";  
RL Nucleic Acids Res. 11:2367-2380(1983).  
CC -1- FUNCTION: SEED STORAGE PROTEIN.  
CC -1- SUBCELLULAR LOCATION: CYTLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN  
CC BODIES.  
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,  
CC CONVICTILIN, CONGLICININ, ETC.).  
CC PIR: A03344; FWPMBV.  
DR HSSP: P50477; ICAM.  
DR INTERPRO: IPR001113; -  
DR PFAM: PF00546; Seedstore\_7s; 1.  
KW Seed storage protein; Glycoprotein; Multigene family; Signal.  
FT SIGNAL 1 15  
FT CHAIN 16 >410 PROVICTILIN.  
FT SITE 221 222 CLEAVAGE (POTENTIAL).  
FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT NON-TER 410 410  
SQ SEQUENCE 410 AA; 46385 MW; 8AF68CE85A316FA2 CRC64;  
  
Query Match 21.4%; Score 712.5; DB 1; Length 410;  
Best Local Similarity 37.9%; Pred. No. 2e-37;  
Matches 151; Conservative 76; Mismatches 142; Indels 29; Gaps 5;  
  
QY 208 DNPYYEDERSLSTRFTEEGHISVLENFYGRSKLRALKNYRLVLLLEANPAFLPHLD 267  
DB 20 ENFIFRSNRFQVLYENENGHILRLQKFDKRSKIFENLQNRLLLEYSKPHPTFLQYND 79  
QY 268 ADAILLVIGRGALKMTHRDRESYNLECGDVIRIPAGTFYLLINRNNRLHIAFLQT 327  
DB 80 ADFLVLVLSGKALITVLLPDRNSFSLERCDITKLPAGTIGLVNRDDEDLRVLDVIP 152  
QY 328 ISIPGOYKEFFPAGGNPEPYLSTFESKLEALINTOTERLRYLQO----- 376  
DB 140 VNRGOLQSFLLSGTONQKSSLSGFSKNILEAFNTYEIEKVLLEQOQDEPOHRSILK 199  
QY 377 -----EGVIRASQEQIRELTRDDESRRMHIRGSGSSRGPNLFRKRPYSNKG 428  
DB 200 DRROGINEENVIVAKISQJLELNKNAKSSS---KSSVSESGPNLRSRNPYSNKG 255  
QY 429 QAVEVKEPEDYROLQDMQVSVFIANTOGSMGPFENRSTKRVVAVASGEADVEMACPRLS 488  
DB 256 KFEFIFPKRNPQLODNLIFVNSVDIKVSLPLPHYNSRAIVTVJTEGKDFELVQGR-- 313  
QY 489 GRHGGGCGKRHEEEVEV---HYQVVARLSKREAIIVLAGHPVYVSSGNENLLFAFGI 546  
DB 314 NENQKENDKEEEOEETSQOVOLYRAKLSPGDVIVIPAGHPVAINNSDNLIL--GLGI 371  
QY 547 NAONHNFENLAGRENVLOQIEPQAMELAFASRKEVE 584  
DB 372 MAENNERNFLAGEEDNVISQVERPVYELAFPGSSHEVD 409  
  
RESULT 14  
CANALCANEN STANDARD; PRT; 445 AA.  
AC P50477;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE CANAVALIN PRECURSOR.  
 OS Canavalia ensiformis (Jack bean) (Horse bean).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
 OC Fabales; Fabaceae; Papilionoideae; Canavalia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=COTYLEDON;  
 RA MEDLINE=92119225; PubMed=1731967;  
 RX Ng J.D., Stinchcombe T., Ko T.-P., Alexander E., McPherson A.;  
 RT "PCR cloning of the full-length cDNA for the seed protein canavalin  
 from the jack bean plant, Canavalia ensiformis.";  
 RL Plant Mol. Biol. 18:147-149(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=COTYLEDON;  
 RX MEDLINE=94143475; PubMed=8310055;  
 RA Ng J.D., Ko T.-P., McPherson A.;  
 RT "Cloning, expression, and crystallization of jack bean (Canavalia  
 ensiformis) canavalin.";  
 RL Plant Physiol. 101:713-728(1993).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).  
 RX MEDLINE=94143476; PubMed=8310056;  
 RA Ko T.-P., Ng J.D., McPherson A.;  
 RT "The three-dimensional structure of canavalin from jack bean  
 (Canavalia ensiformis).";  
 RL Plant Physiol. 101:729-744(1993).  
 CC -1- FUNCTION: SEED STORAGE PROTEIN.  
 CC -1- SUBUNIT: HOMOTRIMER.  
 CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,  
 CONVICILIN, CONGLYCININ, ETC.).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X59467; CAA2075.1; -  
 DR PDB: 1CAU; 31-OCT-93.  
 DR PDB: 1CAV; 31-OCT-93.  
 DR PDB: 1CAM; 31-OCT-93.  
 DR PDB: 1CAX; 31-OCT-93.  
 DR INTERPRO: IPR001113; -  
 DR PFM: PF00546; Seedstore\_7s; 1.  
 KM Seed storage protein; Signal; 3D-structure.  
 FT SIGNAL 26  
 FT CHAIN 1  
 FT SIGNAL 27  
 FT CHAIN 1  
 SQ SEQUENCE 445 AA; 50326 MW; 30383C5F83A1E9B7 CRC64;

Query Match 20.8%; Score 692.5; DB 1; Length 445;  
 Best Local Similarity 33.8%; Pred. No. 3; 9e-36;  
 Matches 144; Conservative 99; Mismatches 160; Indels 23; Gaps 5;

QY 194 GSGRVEEESKSDNPYVYDERLSLSTRTEGHISVLENYGSKLRALKNVVLL 253  
 DB 32 GGEAEDESESRANPNPYLFRSNKFTLFKNQHGSLRLQRENEDETEKLEMLADRVLEY 91  
 QY 254 EANPNFVLPHTLHDAALLVIGRGALMKIHRDNRESYALCGDVIIRIPACTFFYLNR 313  
 DB 92 GSKPMTLLPHHSDDLVLVLEGAILVLPDGRDITKKLOGDAIKIQAGTFYLLNP 151  
 QY 314 DNNERLHIAKPLQITISTPGQYKEFPFAGQNPPEVYLSFSSKELLEALNTQTERL-RGVL 372  
 DB 152 DNNQNLRIKFAITFRPCTVEDFELSTKRLPSYLSAFSKNLEASYSDPYDEIQTLL 211  
 QY 373 GQGRGVITIRASQEQIRELTRDSESRHRIIRGSSSGPYNLPKRLYLNKRYQAAE 432  
 DB 212 QEOEGVIVKMKRQDIQETSKHAOSSSKRTL-----SSQDKPFLNRDPIYSNNYKIVE 267

QY 433 VKPEYRQLQDMQDVSLANITQGSMMCPFFNTIRSTKVVVAVSGEADYBMACPHLSGRHG 492  
 DB 268 IPEKNSQLRDDLILNLCQMGEGALFVPHYNSRAVILVAVNGREVELY----- 318  
 QY 493 GRCGRKHEEEVEHYEYR---ARLSKREALVYLAGHPYVVSNGENLLPARGINAO 549  
 DB 319 ---GLEQDQOOGLESMQRLRYAATLSECDIIVIPSSFPVALKAASDLNMV--GICVNAE 372  
 QY 550 NNNENLAGRRNVLCQIEPQAMETAFASRKEVELEFNSODESJFFPGPROHQOOSPRS 609  
 DB 373 NNERNFLAGHKENVIRQIPROVSDLTFFSGGEVEELLEKNKESYFVDQPHRIDAGKA 432  
 QY 610 TKQOP 615  
 DB 433 RAHLR 438

RESULT 15  
 ID CANA\_CANGL STANDARD; PRT: 445 AA.  
 AC P10562;  
 DT 01-JUL-1989 (rel. 11, created)  
 DT 01-JUL-1989 (rel. 11, last sequence update)  
 DT 01-OCT-1996 (rel. 34, last annotation update)  
 DE CANAVALIN PRECURSOR.  
 OS Canavalia gladiata (Sword bean) (Japanese jack bean).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
 OC Fabales; Fabaceae; Papilionoideae; Canavalia.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 27-40.  
 RC TISSUE=SEED;  
 RX MEDLINE=8811636; PubMed=3338449;  
 RA Yamuchi D., Nakamura K., Asahi T., Minamikawa T.;  
 RT "CDNAs for canavalin and concanavalin A from Canavalia gladiata  
 seeds. Nucleotide sequence of cDNA for canavalin and RNA blot  
 analysis of canavalin and concanavalin A mRNAs in developing seeds.";  
 RL Eur. J. Biochem. 170:515-520(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=SEED;  
 RX MEDLINE=89296493; PubMed=2740227;  
 RA Takei Y., Yamuchi D., Minamikawa T.;  
 RT "Nucleotide sequence of the canavalin gene from Canavalia gladiata  
 seeds.";  
 RL Nucleic Acids Res. 17:4381-4381(1989).  
 CC -1- FUNCTION: SEED STORAGE PROTEIN.  
 CC -1- SUBUNIT: HOMOTRIMER.  
 CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,  
 CONVICILIN, CONGLYCININ, ETC.).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X06733; CAA29910.1; -  
 DR EMBL: X15076; CAA33172.1; -  
 DR PIR: S00281; S00281.  
 DR HSSP: P50477; ICAM.  
 DR INTERPRO: IPR001113; -  
 DR PFM: PF00546; Seedstore\_7s; 1.  
 KM Seed storage protein; Signal.  
 FT SIGNAL 26  
 FT CHAIN 1  
 FT SIGNAL 27  
 FT CHAIN 1  
 FT CHAIN 27  
 FT CHAIN 161  
 FT CHAIN 161  
 SQ SEQUENCE 445 AA; 50289 MW; 739B59CC2292EA80 CRC64;



